A Conceptual Model for Evaluating System Effects of Response to Intervention

Edward S. Shapiro and Nathan H. Clemens

Assessment for Effective Intervention 2009; 35; 3 originally published online Apr 7, 2009;
DOI: 10.1177/1534508408330080

Article

Assessment for Effective Intervention
Volume 35 Number 1
December 2009 3-16

10.1177/1534508408330080
http://aei.sagepub.com
hosted at
http://online.sagepub.com

A Conceptual Model for Evaluating System Effects of Response to Intervention

Edward S. Shapiro
Nathan H. Clemens
Lehigh University, Bethlehem, Pennsylvania

Implementing a Response to Intervention (RTI) system could improve overall student achievement and the way in which students with disabilities are identified. In order to evaluate the effectiveness of an RTI system (i.e., "Is our RTI system accomplishing its stated goals?"), a set of data-based indicators are needed. This paper will describe a set of five measureable indicators from three domains of evaluation that schools can use to obtain frequent feedback on the impact of their RTI system on reading instruction and achievement. The evaluation methodology provides multiple, sensitive metrics that can be used soon after RTI implementation begins, and does not require that schools wait for more long-term, singular outcome measures such as performance on high-stakes tests to determine if the RTI system is functioning as intended. The data used for each indicator and the way in which the data can impact decisions is described. Issues related to RTI evaluation and areas of further research are discussed.

Keywords: *response to intervention; RTI; evaluation*

Response to intervention (RTI) is a methodology aimed at evaluating and addressing the educational needs of all students and identifying students needing intervention beyond what is provided during typical classroom instruction. The RTI process generally involves systematic screening to identify students at risk for not meeting educational goals and implementing increasingly intensive instructional tiers to provide additional or more intensive research-based instruction to students when needed (Batsche et al., 2006). Student RTI is monitored to determine if more intensive intervention is necessary for that student to meet educational goals. RTI holds promise for addressing several key outcomes, including improving overall student achievement (Ervin, Schaughency, Goodman, McGlinchey, & Matthews, 2006); early detection and intervention of students at risk for academic problems (Al Otaiba & Torgesen, 2007); and improving the way in which students are determined to be eligible for special education services, particularly identification students with specific learning disabilities (SLD) (Speece & Case, 2001). A lack of satisfactory response to intensive, research-based interventions implemented with fidelity allows one to be more confident that the student's academic deficits are due more to an underlying disability and less because of inadequate prior instruction (Fletcher, Coulter, Reschly, & Vaughn, 2004). Thus, RTI holds promise for improving the "accuracy" of student referrals for special education evaluation.

Given the provisions under the 2004 reauthorization of the Individuals With Disabilities Education Act that schools may adopt an RTI process for identification of students in need of special education services, school systems have begun to transition toward a tiered instructional system (Boyer, Cook, Palenchar, & Smith, 2007; Cox & Allen, 2007). As schools engage in this process, there is a clear need to evaluate the impact or success of the tiered instructional model at accomplishing the key desired outcomes. The important question that must be asked by stakeholders in implementing an RTI system is, "Is the model accomplishing the goals it was intended to accomplish?" To conduct such an evaluation, a set of measurable indicators on which to evaluate progress is needed. The purpose of this article is to provide a framework for the evaluation of the success of tiered instruction models at the level of schools using several measurable indicators.

Fixson, Naoom, Blasé, Friedman, and Wallace (2005) identified the important stages in the implementation process that move from exploration and adoption to

Authors' Note: This article was completed with partial support from Grant No. 326M050001, Project MP3: Monitoring Progress in Pennsylvania Pupils, from the U.S. Department of Education to the first author. The opinions expressed in this article do not reflect the views of the U.S. Department of Education. Correspondence regarding this article can be sent to Dr. Edward S. Shapiro, Center for Promoting Research to Practice, Lehigh University, L-111 Research Dr., Bethlehem, PA 18015; e-mail: ed.shapiro@lehigh.edu.

program installation, initial implementation, full operation, innovation, and sustainability. The conceptual model of RTI evaluation discussed in this article is focused primarily at the stage of initial implementation. As defined by Fixsen et al. (2005), this stage of implementation requires real change for an organization and often results in "an environment full of personnel rules, social stressors, union stewards, anxious administrators, political pressures, interprofessional rivalry, staff turnover, and diamond-hard inertia" (Fischer, 1983, p. 249).

Published reports of existing RTI models vary in terms of the outcome variables reported. Some studies and reports have used the number of students referred for special education evaluation and the number of special education placements as a measure of outcome evaluation (e.g., Callender, 2007; Hartman & Fay, 1996; Marston, Muyskens, Lau, & Canter, 2003; McNamara & Hollinger, 1997; VanDerHeyden, Witt, & Gilbertson, 2007). For example, VanDerHeyden et al. (2007) evaluated the effectiveness of an RTI model on five dimensions, each primarily focused on evaluations for eligibility for special education services. They evaluated the impact of the model on the number of evaluations for special education services and the percentage of evaluations that qualified for services, the degree to which school evaluation teams utilized data from the RTI process in determining whether an evaluation should be conducted, the effect of the RTI implementation on rates of identification with regard to ethnicity and gender, and whether the implementation of the RTI model reduced assessment and placement costs for the school district.

Other studies have reported the number of students in a given population requiring additional intervention and the number of those students who responded to the intervention or did not respond and required additional or more intensive intervention (e.g., Ardoin, Witt, Connell, & Koenig, 2005; McMaster, Fuchs, Fuchs, & Compton, 2005; Vaughn, Linan-Thompson, & Hickman, 2003; Vaughn, Wanzek, Linan-Thompson, & Murray, 2007). Some studies have also reported change in student achievement levels or growth in academic skills following the implementation of an RTI model (e.g., L. S. Fuchs, Fuchs, & Prentice, 2004; Gettinger & Stoiber, 2008; Marston et al., 2003; McMaster et al., 2005).

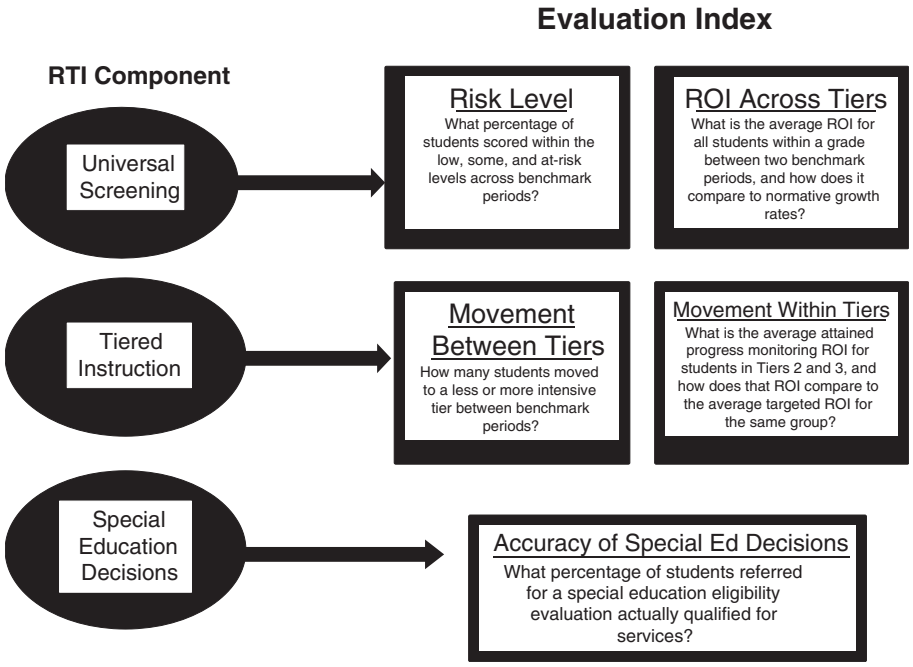
Peterson, Prasse, Shinn, and Swerdlik (2007), in their description of an RTI service delivery model being implemented in Illinois, posed several evaluation methods of the program: (a) whether the model is effective in meeting the needs of students, as measured by positive student outcomes (teacher ratings of student academic improvement, the percentage of students who met their individualized goals, and mean increase in words read correctly per minute on oral reading fluency probes); timeliness of interventions (the average number of days between a student referral for problem solving and the start of an intervention); and the number of reading goals developed for students; (b) the effect of the model on resources and services for students eligible for special education, that is, whether the implementation of the model would adversely affect the delivery of special education services and the number of individuals available to implement the interventions; (c) the effect of the model on the timeliness of referral, evaluation, and service delivery for students suspected of having a disability and any change in the roles of school psychologists and social workers under the model; and (d) whether parents and school staff were satisfied with the model, as measured by surveys with school personnel and parents on their satisfaction with the model as well as surveys with parents on their level of involvement in the problem-solving process.

Bollman, Silberglitt, and Gibbons (2007) cited three outcomes as evidence of the effectiveness of the RTI model implemented by the St. Croix River Education District in Minnesota: (a) an increasing percentage of students across the district who met spring benchmark targets on curriculum-based measurement literacy measures across 10 years, (b) an increasing percentage of students across the district who met the grade-level standard on statewide standardized achievement tests across 10 years, and (c) a reduction in the incidence of learning disabilities across 10 years compared to the incidence observed across the state.

Broad, overarching goals such as improving overall student achievement and improving the process of identifying and placing students into special education are central outcomes of an RTI process. Evaluation of these broad outcomes is critical in ultimately determining if an RTI model is effective. However, changes on these dimensions take time and require a model to be in effect across a full year or even several years to see the full impact on these evaluation dimensions. Thus, a set of short-term, sensitive indicators is needed to evaluate the effectiveness of the initial implementation of an RTI model.

The purpose of this article is to present a conceptual model that will propose a set of five indicators that can enable schools to measure the impact of their RTI system, some of which may provide information after only a few months of implementation. As illustrated in Figure 1, each of the indicators is positioned as a major outcome from each of three key components of the RTI process: universal screening, tiered instruction, and special education decision making. At the universal screening component

Figure 1
Conceptual Model for Evaluating a Response to Intervention (RTI) Model



Note: ROI = rate of improvement.

of RTI, data are used to identify the particular tier into which students fall based on their current level of skill development. Both the impact on changes of student level of risk for successful future academic performance and the rate of improvement (ROI) over time on the universal screening measures serve as critical indices of success. During tiered instruction, indices that measure outcomes involve the degree to which students progress from tier to tier as well as their ROI while in tiered instruction. Finally, the accuracy of decision making related to special education eligibility is another index of RTI outcomes. Each index is posed in the form of an evaluation question, and the collective response across these questions offers a full picture of the impact that the RTI model is having on student outcomes.

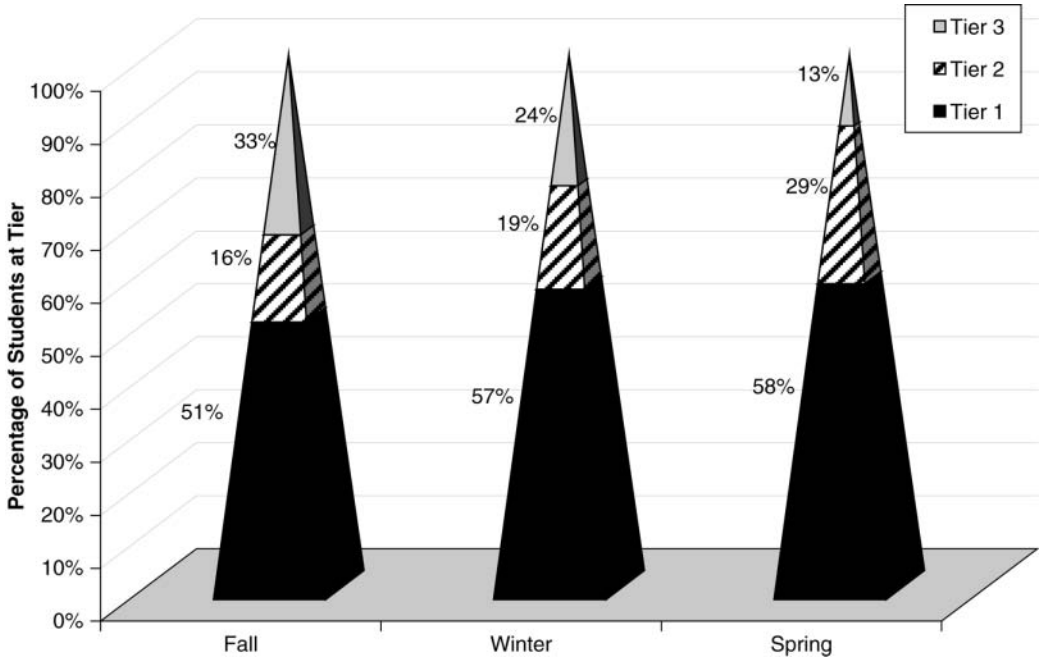
The emphasis of the conceptual model is on the evaluation of the RTI process at the level of grades, schools, and/or districts. Although determining whether individual students are responding to a particular intervention is important, the evaluation strategies for addressing the larger question of whether an RTI model is achieving its desired impact relies on the aggregation of data across individual students. Evaluating aggregated data across grades, schools, or districts allows for teams and administrators to make systemic data-based decisions to improve or enhance their RTI system.

Evaluation Index 1: Monitoring Risk Levels Across Benchmark Periods

Key Evaluation Question: What percentage of students scored within the low-, some-, and at-risk levels across benchmark periods?

Models of RTI are all based on the concepts of identifying students early in their school careers who may be at risk for academic failure. The process involves the administration of universal screening measures whose cut points are set to effectively sort students into categories of relative risk. Those whose performance is at or above the cut point are predicted to have high probabilities of future success in the domain of assessment. Those falling below the cut point are viewed as having high probabilities of subsequent challenge and struggles. Although the selected cut points will always have false positives and false negatives, the use of empirically derived cut points that have demonstrated relationships to desired outcomes on criterion measures (i.e., statewide assessment, performance at subsequent grades) will produce the smallest number of these incorrect predicted decisions (Good, Simmons, & Kame'enui, 2001; Shapiro, Solari, & Petscher, 2008).

Figure 2
Change in Tier Placement by Assessment Period in Third Grade of a School Implementing a Response to Intervention (RTI) Model

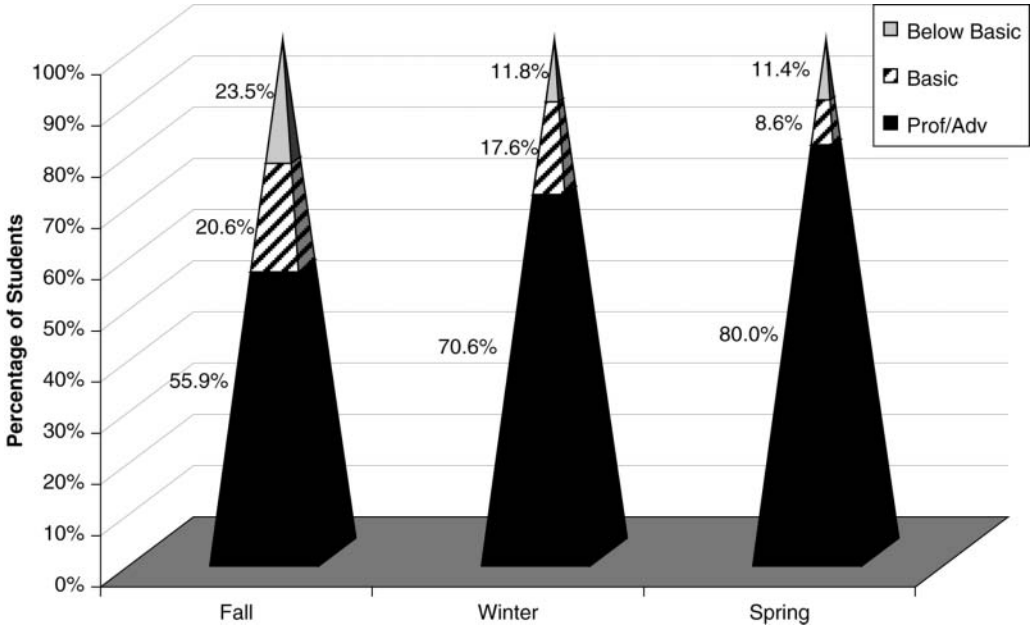


Logically and theoretically, if an RTI model is successful, one expects to move more students toward low risk and reduce the at-risk category. Students would move in the direction of some risk to low risk (Tier 2 to Tier 1) or at-risk to some risk (Tier 3 to Tier 2). Analysis of risk-level movement across the year, comparing the percentage of students moving toward the low-risk category versus those moving toward the at-risk category, would be a reflection of the degree to which the RTI model is "working."

In the area of reading, which has been the focal point of many RTI implementation models reported in the literature, the primary measure used to reflect student performance has been oral reading fluency (ORF) beginning with the end of Grade 1. Known as a general outcome measure (GOM) of reading, ORF has a long and well-established history as a key indicator of overall reading achievement (L. S. Fuchs, Fuchs, Hosp, & Jenkins, 2001; Shinn, Good, Knutson, Tilly, & Collins, 1992). Indeed, there are several decades of research substantiating the strong correlations of ORF with many other broader measures of reading performance such as published norm-referenced standardized tests (e.g., Deno, Mirkin, & Chiang, 1982; L. S. Fuchs, Fuchs, & Maxwell, 1988; Shinn et al., 1992) and statewide high-stakes assessments (Good et al., 2001; Hintze & Silberglitt, 2005; McGlinchy & Hixson, 2004; Shapiro, Keller, Lutz, Santoro, & Hintze, 2006; Stage & Jacobsen, 2001). Furthermore, ORF has been shown to have long-term predictability as a measure of reading across 1- and 2-year-later outcomes (Baker et al., 2008; Good et al., 2001; Hintze & Silberglitt, 2005; Keller-Margolis, Shapiro, & Hintze, 2008). Certainly, risk levels as reflected on this powerful measure of overall reading performance are important indicators of RTI impact.

What we expect to see is a pattern illustrated in Figure 2. These data as well as data presented in all other figures come from the actual implementation of an RTI model. Figure 2 displays the risk outcomes for students in Grade 3 on ORF from a school that had implemented an RTI model. Looking at the figure, one can see that during the year, the percentage of students found on the ORF measure improved from 51% in fall to 58% in spring. In this particular grade, there was a reduction from 33% of students at risk in fall to 13% at the end of the spring. Increasing numbers of students at the Tier 1 level with decreasing numbers of students at the Tier 3 level is viewed as a favorable outcome and evidence that the model was having the desired impact. Clearly, the impact of this school's model of RTI on ORF was focused strongly on students identified as at risk and moving those students to categories of less risk. Less impact was found

Figure 3
Levels of Performance on *4Sight Benchmark Assessment* Measure of Reading for Grade 3 of a School Implementing a Response to Intervention (RTI) Model



in moving students from the some-risk to the low-risk category, at least as reflected on the ORF measure.

Although risk level related to ORF in analyzing RTI models is certainly important, one also needs to look beyond ORF as additional metrics can enhance diagnostic decision making, especially for students in Grade 3 and later. Shapiro et al. (2008) and Clemens and Shapiro (2008) demonstrated that for older elementary students, the decisions can be enhanced by adding data from a measure of reading comprehension in addition to ORF data. Given that ORF can begin to lose its sensitivity to reflect outcomes for older elementary and middle school students (Catts, Hogan, & Adlof, 2005; Gough & Tunmer, 1986; Silberglitt, Burns, Madyun, & Lail, 2006), the risk levels reflected through measures beyond ORF also need to be considered when examining the impact of RTI models. Figure 3 shows the risk levels on a repeated measure of reading comprehension (*4Sight Benchmark Assessment*; Success for All Foundation, 2008) administered three times per year for students in Grade 3 to the same group of students as Figure 2. The data in Figure 3 reflect a far greater number of students scored in the low-risk category on the measure of reading comprehension than those on the ORF measure. For example, at spring, the outcomes of the data collected from the ORF measure showed that 29% of students would be viewed as at some risk or Tier 2. By contrast, data from Figure 3 on the *4Sight Benchmark Assessment*, a measure that is highly correlated to the statewide assessment (Shapiro et al., 2008), would have shown 8.6% of students at some risk or Tier 2. Comparing Figures 2 and 3 shows that the risk level reflected in ORF alone may have over identified students who are at risk. Typically, the overidentification process occurs because cut scores on ORF measures are designed to contain more false positives, thereby making sure that all students with any chance of failing are provided with the needed resources to succeed.

Change in the percentage of students at risk is one key indicator of the impact of RTI models. Ideally, one would want to see the largest proportion of students within the low-risk or Tier 1 level. The proposed targets of 80% of students falling within a low-risk category, 15% of students at a category of some risk, and 5% of students falling within an at-risk category and in need of substantial intervention (e.g., Batsche et al., 2005) are akin to the recommendations from School Wide Positive Behavior Support, which in turn was based on a public health model of prevention (Walker et al., 1996). However, districts with a student population that is highly transient or beset with risk factors that impact educational achievement may find it difficult, if not

impossible, to ever achieve the hypothetical 80% level of students at a low-risk category. In this case, a goal would be to strive for getting the majority of students to a low-risk level and trying to increase that percentage over time. Additionally, one could empirically derive the level of risk that best predicts to a key criterion, for example, to state achievement test scores or some other nationally standardized achievement test, specifically for the local school or district. This would give a truer picture of the risk levels within the local context and serve as a better barometer for measuring the degree to which outcomes of low risk should be established.

Evaluation Index 2: Rate of Improvement (ROI) Across Benchmark Measures

Key Evaluation Questions: What is the average ROI for all students in a grade between two benchmark periods, and how does that ROI compare to normative growth rates and/or previously observed growth rates?

A second way to see if RTI models are "working" is to look at the overall ROI of students between benchmark periods. Universal screening measures collected at three points in time provide data that show how students are changing over time. These data can be aggregated at the level of an entire grade, school, or district. Normative data available from national databases can identify the rate of progress that would be expected of typical performing students. For example, in reading, the DIBELS (Dynamic Indicators of Basic Early Literacy Skills) (Good et al., 2001) identified the predicted benchmark scores at each point in the universal screening that would indicate with 90% probability a student's success in the subsequent year. Using conditional probability based on the Oregon Reading First sample, the developers of the DIBELS provided scores that would indicate the bottom of the low-risk category. By examining the performance of students at fall, winter, and spring, one can determine the expected rate of growth for typical performers for that grade. Other developers such as AIMSweb (Pearson Education, Inc., 2008) have provided large normative databases across multiple districts that offer similar data indicating the rate of improvement observed across students starting within different percentile ranks. Moreover, empirical studies conducted across several years have established normative levels of achievement in ORF from which expected rates of improvement can be derived (Hasbrouck & Tindal, 1992, 2005).

Using the ROI of typically performing students, one can compare the ROI of the grade, school, or district to determine the degree to which those in RTI models are performing similar to those in the national normative databases. Table 1 offers an example of two groups of Grade 2 students in two different elementary schools implementing an RTI model. The average scores for students in the fall, winter, and spring of the school year is displayed along with the average ROI of each group based on these scores compared to the ROI expected of students meeting the DIBELS benchmarks across the school year. As can be seen in Table 1, the second-grade students in School 1 achieved average scores in the fall, winter, and spring that were slightly higher than targeted for those points in time during the year, but they demonstrated an average ROI equal to that expected of students meeting the DIBELS benchmarks across the year. Students in School 2 showed a similar starting point to School 1 at the fall of the year but grew at a rate somewhat greater by spring of the year. Analyzing the ROI of students across the tiers provides an index of the rate at which the group of students is progressing, and comparing that rate to normative or expected ROIs allows educators to determine whether students are progressing at rates commensurate with what is expected of students at a given grade level. When a group of students is observed to demonstrate an ROI that exceeds that which is typically observed or expected, it can be taken as evidence that the current instruction is succeeding in helping students to progress at an ideal rate. This type of analysis can add important information to what is found through Evaluation Index 1. As mentioned previously, districts with historically low-performing student populations may not see many students move to a low-risk category between two or three benchmark periods. However, by looking at the average ROI of a group, one can determine if the students in the group are "gaining ground," thus making them more likely to achieve low-risk targets in the future.

Table 1
Rate of Growth on Benchmark Assessments of Two Schools Across the Year

	Fall	Winter	Spring	Growth Rate
DIBELS benchmarks:	44.0	68.0	90.0	1.4 WRC/week
School 1: Grade 2	48.0	75.4	95.2	1.4 WRC/week
School 2: Grade 2	47.6	77.6	99.3	1.5 WRC/week

Note: DIBELS—Dynamic Indicators of Early Literacy Skills; WRC—Words Read Correct.

A cautionary note about the use of ROI must be offered when used at the individual student level. Because ROI is based on the establishment of a trend from a series of data points, trend is sensitive to many

factors including the number of data points used to establish the trend. In looking at ROI based on benchmark data, one would be drawing conclusions using only two or three data points. Trends derived from so few data points would be greatly influenced by a single data point's deviation. However, by aggregating the data across individuals within an entire grade, the impact of individual variation would be lessened. Another important point to consider in the use of slope as an indicator of outcomes is that slope is impacted by the intercept or beginning of the year starting point for the student's performance (Silberglitt & Hintze, 2007; VanDerHeyden, Witt, & Barnett, 2005). Clearly, students who begin at a lower level and grow during the year would have very different predicted outcomes than those who may start just below benchmark but do not change much during the course of the year.

Schools that have been collecting universal screening data for more than one year have an opportunity for a deeper level of analysis. Schools with multiyear data can compare the average ROI observed for the current year to the ROI observed in previous years. In this way, schools can evaluate whether changes to instruction have resulted in increasing the rate of student progress from year to year.

Evaluation Index 3: Movement Between Tiers

Key Evaluation Question: How many students moved to a less or more intensive tier between benchmark periods?

Tiered intervention is a core feature of any RTI model, as the purpose of the model is to accelerate the performance of students whose trajectories predict learning outcomes that are not likely to reach benchmark levels of performance. As the year progresses, students who are responding to interventions are expected to reach levels that will allow movement from more intensive to less intensive tiers. Those students assigned to Tier 3 or Tier 2, if progress is indeed made as expected, would be moved back to Tier 2 or Tier 1, as appropriate. Likewise, students who are not making the expected progress through interventions may need even more intensive levels of supplemental instruction and are therefore moved from Tier 1 or 2 to Tier 2 to 3. Movement between the tiers is a reflection of the impact of an RTI model. When data indicate that greater movement occurred from more intensive to less intensive tiers, it suggests the RTI model is succeeding in helping students make gains that allow them to be placed at less intensive tiers. On the other hand, when data indicate that the majority of students are moving toward more intensive intervention tiers, it would suggest that the model should be reexamined with regard to the instructional effectiveness of the Tier 1 (core instruction) and Tier 2 intervention. Tier movement data obtained through this index is a reflection of team decision making and agreement that students indeed are in need of more or less intensive tiered levels of instruction.

Figure 4
Aggregated Data of Tier Movement for Four Schools in the Same District Across One Year Within a Response to Intervention (RTI) Model

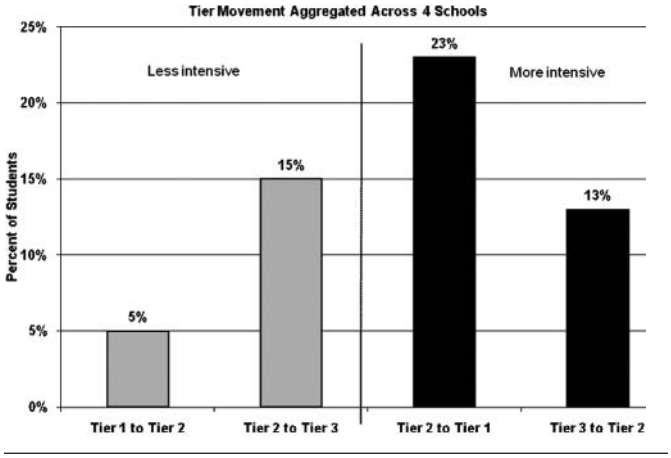


Figure 4 illustrates the impact of tier movement across four schools implementing similar RTI models. Each school had three defined tiers of intervention, each tier discriminated by the nature of instruction as well as frequency of supplemental intervention. Each school assigned students to tiers on the basis of universal screening. Progress monitoring was collected for students at Tier 2 every other week and once a week at Tier 3. Meeting monthly as grade-level teams, and quarterly as schoolwide data analysis teams, the school staff looked at student progress and made recommended changes in intervention groups as well as tier assignment.

Aggregated across grades and across schools, the data reflect that across the year, 36% of students moved from more intensive to less intensive tiers (i.e., Tier 3 to 2 or Tier 2 to 1), and 20% moved to more intensive instructional tiers (from Tier 1 to 2 or Tier 2 to 3). As shown in Figure 4, there was a net gain of 16% of students moving to less intensive tiers compared to more intensive tiers. Again, this adds another dimension of evaluation of the impact of the RTI model.

For a more specific evaluation of the tier movement that occurred at each grade level, the same analysis can be repeated using data from a specific grade. Separate grade-level analyses such as this allow for teams to evaluate the effectiveness (or lack thereof) of intervention programs at specific grade levels. An interesting and as-yet-unevaluated aspect of tier movement is the degree to which team decision making is consistent with the recommendations that are derived from the data collection process. In other words, although individual student universal-screening or progress-monitoring data may suggest that a change in tiered intervention would be logical, the ultimate decision of moving students to more or less intensive tiered intervention is a function of team decision making. Research is needed to examine the degree of congruence between data-based recommendations and team decision making within RTI models. The degree to which congruence changes over time as schools become more comfortable with RTI decision-making processes is also of importance in fully evaluating the RTI model.

Evaluation Index 4: Movement Within Tiers

Key Evaluation Questions: What is the average attained ROI on progress-monitoring measures for students in Tiers 2 and/or 3, and how does that ROI compare to the average targeted ROI for the same group?

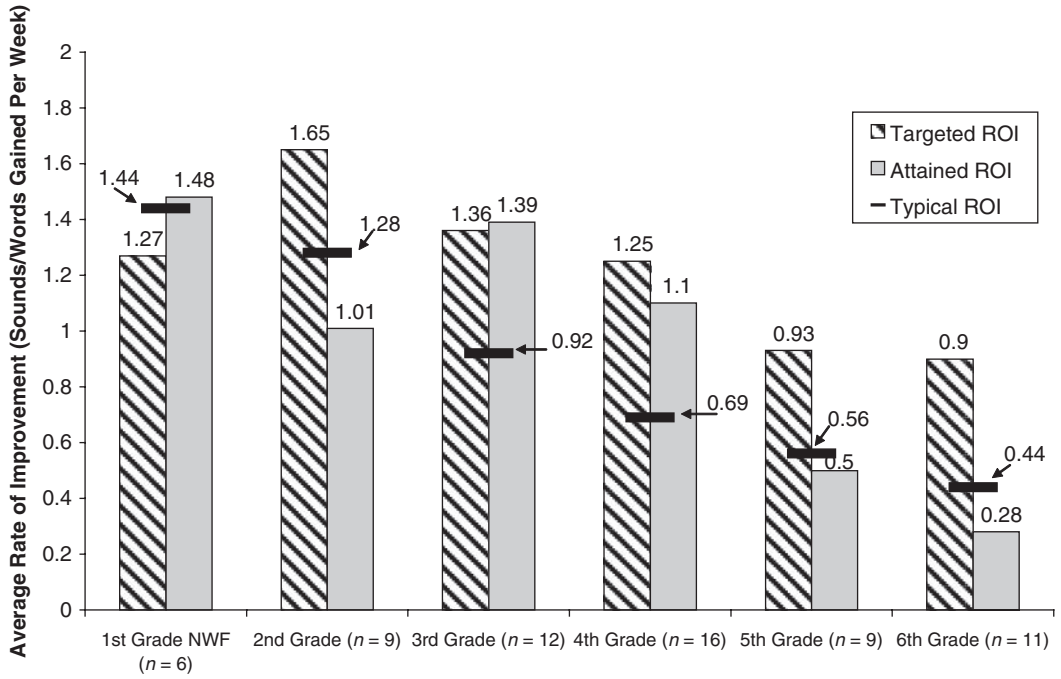
When students are assigned to Tier 2 or 3 instruction, progress monitoring of student performance is conducted to provide more frequent data on the impact of supplemental instruction. The frequency of progress monitoring is on a schedule that matches the intensity of the instructional tier. Those at Tier 2 are usually assessed at least monthly (although we personally recommend a schedule of no less than every 2 weeks). Those at Tier 3 are monitored at least once a week. Three key indicators must be considered in evaluating movement within tiers: *typical* ROI, or the average ROI observed in the general population based on normative data; *target* ROI, the ROI targeted for an individual student; and *attained* ROI, or the ROI a student or group of students actually achieved.

For students in Tier 2 or 3, individual goals are established to determine if students are responding to the intervention in ways that would result in their meeting performance objectives of those students currently at Tier 1. Given that all students assigned to tiers would by definition be below benchmark levels through the universal screening process, and that goals for these students would be those that reflect a "closing" of the achievement gap, all of these students should have an individual targeted rate of progress that would exceed the rate expected of typical performing students. For example, a second-grade student, Chris, achieving the benchmark level performance on an ORF measure at the beginning, middle, and end of the year, would be expected to go from 60 words correct per minute (wcpm) to 110 by the end of the year. Assuming a 36-week school year, Chris would have a benchmark ROI of 1.38 wcpm per week. (110 – 60 = 50/36 = 1.38). Consider a second student, Sam, identified as in need of Tier 2 instruction with a fall ORF score of 50. Setting the goal for Sam to reach the benchmark level of 110 wcpm by the end of the year would mean that his targeted ROI would be 1.67 wcpm per week (110 – 50/36 = 1.67). In other words, Sam needs to move at a rate greater than that of a typically performing student to close the gap toward the benchmark goal.

Students assigned to tiered instruction will remain in that supplemental instructional program until a clear trend can be established through their progress-monitoring data. It has been recommended that a minimum of 10 data points be collected before evaluating trend (Good & Shinn, 1990). Indeed, longer durations of progress monitoring with more data points will reduce the standard error associated with slope estimates and improve the stability of the slope estimate (Christ, 2006). If progress monitoring is being collected every 2 weeks, this means that at a minimum, students will remain in tiered instruction for 8 to 12 weeks. Looking across these students assigned to tiered instruction, one can aggregate the attained ROI for each student to better determine the degree to which the students are "beating" the typical ROI as well as their own personal targeted ROI. Looking across grades, schools, or districts, one can get another indication of the degree to which an RTI model is working.

Figure 5 reflects an example from a school that had been implementing an RTI model. Looking at Grade 3 for ORF, one can see that the mean targeted ROI across the 12 students in tiered instruction was 1.36 wcpm/week. The attained ROI for these same students was 1.39 wcpm/week. Typically performing Grade 3 students meeting benchmark targets would be expected to achieve an ROI of 0.92 (see the black bar in the figure). These data indicate that in this third-grade sample, students demonstrated an average attained ROI that met the average targeted ROI and exceeded the ROI observed in typically performing students. This information suggested that, at least at third grade, the tiered instruction had been successful in helping students meet individual targets as well as exceed typical rates of improvement (thus allowing them to "close the

Figure 5
Targeted, Attained, and Typical Rate of Improvement (ROI) From Progress-Monitoring Data Across Grades at One School Implementing a Response to Intervention (RTI) Model



Note: NWF = nonsense word fluency; data for other grades are for oral reading fluency.

gap"). The data from other grades were not as positive; in second grade, for example, the average attained ROI was well below both the targeted ROI and the typical ROI. The tiered instruction being delivered at this grade level may need to be improved.

The desirable outcome is that while students are in tiered interventions, they should be making progress that places them on target to reach goals that would equal levels of students at lower tiers. Although their performance may be such that they have not yet reached performance levels that suggest to teams to move the student to a lower tier, the progress they are making toward tier movement plays an important role in understanding the impact and evaluation of RTI models.

As with using ROI with benchmark assessments from Evaluation Index 2, certain cautions must be considered when using ROI for progress monitoring. In addition to the number of data points used to generate ROI, factors such as variability of passage difficulty, standardization of administration, and integrity of administration can all create variability of individual data points which can effect the calculation of slope. When using progress-monitoring ROI, it is critical that teams determine if any of these factors may be having a substantial impact on the attained data.

Evaluation Index 5: Accuracy of Referrals to Special Education

Key Evaluation Question: What percentage of students referred for a special education eligibility evaluation actually qualified for services?

A final indicator of the impact of an RTI model revolves around the frequency and outcomes of referrals to special education. RTI defines students in need of special education as those who have not responded at the desired level and rate given successively more intensive levels of supplemental instruction within the general education program (i.e., Burns & Senesac, 2005; D. Fuchs, Fuchs, McMaster, & Al Otaiba, 2003). Although RTI models may differ somewhat, with some models specifying one or two levels of supplemental intervention prior to special education (e.g., Batsche et al., 2006; D. Fuchs, Mock, Morgan, & Young, 2003), students who have been provided with supplemental interventions at Tier 2 or 3, and are not making progress at a level viewed as likely to reach educational outcomes that can be maintained within the general education setting, may be in need of special education services.

Students found eligible for special education services will require an individual educational plan (IEP) that indicates the nature and need for accommodations throughout the school life, cutting across content and academic areas. They may need accommodations related to test taking, modification in time limits for assignments, and other such typical changes in their educational program to be successful. Likewise, their academic levels must be by definition substantially below age/grade level expectations such that progress monitoring may need to be adjusted to instructional rather than grade level. In other words, although progress monitoring is typically conducted with grade-level material to determine student performance toward grade-level expectations, monitoring at grade level initially makes little sense when a student's instructional level is substantially lower than his or her chronological grade. As students improve their performance and approach their chronological grade level, progress monitoring at grade level would become a better indicator of outcome of their instructional program.

An important potential outcome of RTI would be that for the students brought to the level of a comprehensive evaluation for eligibility, a very high percentage would be found as eligible for special education. In other words, the efficiency ratio of students evaluated to those found eligible would be greatly increased by identifying the students with an underlying disability rather than students who have not succeeded due to inadequate instruction. Consider a school in which, in a typical year, a multidisciplinary evaluation team has 50 students referred for evaluations, with 35 of those students actually found eligible for special education services (70%). If, in the next year, improvements are made to the prereferral process and the team now has 25 referrals for evaluation, with 23 of 25 students (92%) found eligible, it would represent an improvement in the efficiency of the referral process. VanDerHeyden et al. (2007) found that following the implementation of an RTI model across five schools, the percentage of students found eligible for special education increased in four out of five schools. In one school, for example, the percentage improved from 41% in the first year (no RTI) to 71% the following year after RTI was implemented. The improvement in efficiency is extremely important for those who are required to conduct the comprehensive evaluation, which is a very expensive and time-consuming part of the process of placing students into special education. Indeed, the data obtained from the RTI process for a student being considered for special education very much can be a significant part of the required comprehensive evaluation needed for determinations of eligibility for special education services. From an administrator's perspective, savings are possible of both time and money on evaluation processes that can be better spent on instruction and reallocation of educator roles in support of the RTI process.

Given that RTI is conceptualized as a method of preventing the development of more serious academic problems by intervening early and not allowing students to experience successive failures before intervention is provided (Vaughn, Wanzek, Woodruff, & Linan-Thompson, 2007), it is anticipated that the number of students potentially in need of the intensity of instruction offered through special education will be potentially diminished from current levels. At the same time, the RTI model is designed so that students truly in need of special education services will indeed receive these services. Indeed, early data from longitudinal implementation of RTI have begun to show such outcomes (e.g., Bollman et al., 2007; VanDerHeyden et al., 2007). In these studies, reductions of referrals to special education have been substantial and truly reflect the anticipated outcomes that were expected when RTI was put forward as an alternative to identification of SLD in the Individuals with Disabilities Education Improvement Act of 2004.

Conclusions and Recommendations for Further Research

The conceptual framework presented here can provide a methodology for bringing together data from different sources and components of the RTI process in a way that school personnel can better ascertain the impact of their model. The conceptual framework provides a mechanism to determine if what they are doing is working and, in a formative assessment method, to inform changes needed to the model to enhance its objectives. The model also provides multiple sensitive metrics of progress, as opposed to looking at singular metrics such as outcomes on high-stakes statewide testing. Figure 6 provides a worksheet that schools can use to summarize the key data from the five evaluation questions.

Comprehensive evaluation using the methods described in this article underscores the importance of good data organization and management. To make the evaluation process feasible, timely, and productive, key individuals at the building or district level should be identified who will provide a leadership role in data organization and management. The evaluation methods described do not require the use of any sophisticated statistical software beyond Microsoft Excel (or a similar application), but the individuals involved in data organization and management should have a good working understanding of the program, as well as the use of any assessment management software being used such as AIMSweb or the DIBELS data system.

Figure 6
Summary Data Worksheet for Organizing Evaluation of Response to Intervention (RTI)

Data Organization Worksheet: Indicators of RTI evaluation

School/District _____ Grade _____ School Year _____

1. *What percentage of students scored within the low, some, and at risk levels across benchmark periods?*

	Fall	Winter	Spring	Improvement? F to W	Improvement? W to SP	Comments, observations
% at Tier 1				**Yes ☐**	**Yes ☐**	
% at Tier 2				**Yes ☐**	**Yes ☐**	
% at Tier 3				**Yes ☐**	**Yes ☐**	

2. *What is the average ROI for all students in a grade between two benchmark periods, and how does that ROI compare to normative growth rates and/or previously observed growth rates?* **Measure (e.g., ORF): ____**

	*Fall to Winter Avg. ROI**	*Winter to Spring Avg. ROI**	*Full Year ROI**	Comments, observations
Grade _____				
Comparison:	≥ Norm ☐	≥ Norm ☐	≥ Norm ☐	
_____	< Norm ☐	< Norm ☐	< Norm ☐	

*ROI = (ending score – starting score) / # of weeks

3. *How many students moved to a less or more intensive tier between benchmark periods?*

Fall to Winter Analysis	**# at Tier 1 in Winter**	**# at Tier 2 in Winter**	**# at Tier 3 in Winter**	**Summary**	**Comments, observations**
# of students at Tier 1 in Fall:_____				% stayed at T1:_____ % dropped to T2/3:_____	
# of students at Tier 2 in Fall:_____				% moved up to T1: _____ % dropped to T3:_____	
# of students at Tier 3 in Fall:_____				% moved up to T1/2: ____ % stayed at T3: _____	

Winter to Spring Analysis	**# at Tier 1 in Spring**	**# at Tier 2 in Spring**	**# at Tier 3 in Spring**	**Summary**	**Comments, observations**
# of students at Tier 1 in Winter:_____				% stayed at T1:_____ % dropped to T2/3:_____	
# of students at Tier 2 in Winter:_____				% moved up to T1: _____ % dropped to T3:_____	
# of students at Tier 3 in Winter:_____				% moved up to T1/2: ____ % stayed at T3: _____	

4. *What is the average attained ROI on progress monitoring measures for students in Tiers 2 and/or 3, and how does that ROI compare to the average targeted ROI for the same group?*

	Fall to Winter				**Winter to Spring**		
	Avg. Targeted ROI	**Avg. Attained ROI**	**Comments**		**Avg. Targeted ROI**	**Avg. Attained ROI**	**Comments**
Tier 2 Students				**Tier 2 Students**			
Tier 3 Students				**Tier 3 Students**			

5. *What percentage of students referred for a special education eligibility evaluation qualified for services?*

students referred: _____ # students found eligible for services: _____ "Accuracy" %: _____

Note: ROI = rate of improvement; ORF = oral reading fluency.

The impact of change through an RTI model takes time to fully experience, with many experts indicating a 3- to 5-year time span for full implementation. However, educators cannot wait 3 years to fully decide if their RTI model is effective in meeting goals. The conceptual framework offered here provides a lens by which one can look at the movement toward success during the initial implementation process. In the early stages of an RTI model, one may not see changes in some areas such as risk levels at some grades. Likewise, dramatic changes in tier movement during a year may or may not be evident. But by looking at the multiple metrics during early periods of implementation and then examining these metrics over time, the trends that suggest the model is effective or needs to be altered can be better determined. Most of the indices proposed in this article can be examined after only a few months of RTI implementation and revisited subsequently at later intervals. Additionally, multiyear RTI implementation affords schools the opportunity to use these indices to evaluate improvement on a longitudinal basis. When desirable trends are not observed, educators are provided with evidence for intervention and alteration of model parameters.

Although we have now applied this conceptual model for many RTI implementation models, there is a need for continued empirical substantiation that the model does indeed reflect overall impact of RTI implementation. Other important evaluation indices in addition to the ones proposed here may emerge through research and application of RTI systems. In addition, through the consistent application of the same model across many different RTI implementation efforts in many different contexts, the strength of the model to be a potential mechanism for standardizing the framework for the effectiveness of RTI can be better determined.

Actual implementation of RTI across the country is still in its infancy. As more and more schools begin to adopt RTI models, it will take time and evaluation to fully understand the nuances of RTI and its impact on student outcome variables. The evaluation methodology described in this article provides a starting point for schools and districts to critically and objectively evaluate the effectiveness of their RTI system and make improvements where needed. Over time, the use of continuous, data-based system evaluation improves the ability of that system to effectively meet the needs of all students.

References

Al Otaiba, S., & Torgesen, J. (2007). Effects from intensive standardized kindergarten and first-grade interventions for the prevention of reading difficulties. In S. E. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention* (pp. 212–222). New York: Springer.

Ardoin, S. P., Witt, J. C., Connell, J. E., & Koenig, J. L. (2005). Application of a three-tiered response to intervention model for instructional planning, decision making, and the identification of children in need of services. *Journal of Psychoeducational Assessment*, *23*, 362–380.

Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J. R., Kame'Enui, D. J., et al. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review*, *38*(1), 18–37.

Batsche, G., Elliot, J., Graden, J. L., Grimes, J., Kovaleski, J. F., & Prasse, D. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria, VA: National Association of State Directors of Special Education

Bollman, K. A., Silberglitt, B., & Gibbons, K. A. (2007). The St. Croix River Educational District model: Incorporating systems-level organization and a multi-tiered problem-solving process for intervention delivery. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 319–330). New York: Springer.

Boyer, L., Cook, N., Palenchar, L., & Smith, S. (2007, December). *State presentation: West Virginia*. Paper presented at the 2007 RTI Summit, Arlington, VA.

Burns, M. K., & Senesac, B. V. (2005). Comparison of dual discrepancy criteria to assess response to intervention. *Journal of School Psychology*, *43*, 393–406.

Callender, W. A. (2007). The Idaho results-based model: Implementing response to intervention statewide. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 331–342). New York: Springer.

Catts, H. W., Hogan, T. P., & Adlof, S. M. (2005). Developmental changes in reading and reading disabilities. In H. Catts & A. G. Kamhi (Eds.), *The connections between language and reading disabilities* (pp. 41–54). Hillsdale, NJ: Lawrence Erlbaum.

Christ, T. J. (2006). Short-term estimates of growth using curriculum-based measurement of oral reading fluency: Estimating standard error of the slope to construct confidence intervals. *School Psychology Review*, *35*, 128–133.

Clemens, N. H., & Shapiro, E. S. (2008, August). *Improving diagnostic accuracy in reading within RTI models*. Poster session presented at the annual convention of the American Psychological Association, Boston.

Cox, D., & Allen, M. (2007, December). *State presentation: Virginia*. Paper presented at the 2007 RTI Summit, Arlington, VA.

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children*, *5*, 36–45.

Ervin, R. A., Schaughency, E., Goodman, S. D., McGlinchey, M. T., & Matthews, A. (2006). Merging research and practice agendas to address reading and behavior school-wide. *School Psychology Review*, *35*, 198–223.

Fischer, D. (1983). The going gets tough when we descend from the ivory tower. *Analysis and Intervention in Developmental Disabilities*, *3*(2-3), 249–255.

Fixson, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI Publication No. 231). Tampa: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network.

Fletcher, J. M., Coulter, W. A., Reschly, D. J., & Vaughn, S. (2004). Alternative approaches to the definition and identification of

learning disabilities: Some questions and answers. *Annals of Dyslexia*, *54*, 304–331.

Fuchs, D., Fuchs, L. S., McMaster, K. N., & Al Otaiba, S. (2003). Identifying children at risk for reading failure: Curriculum-based measurement and the dual-discrepancy approach. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 431–449). New York: Guilford.

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definition, evidence, and implications for the learning disabilities construct. *Learning Disabilities: Research and Practice*, *18*, 157–171.

Fuchs, L. S., Fuchs, D., Hosp, M., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*, 239–256.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education*, *9*, 20–28.

Fuchs, L. S., Fuchs, D., & Prentice, K. (2004). Responsiveness to mathematical problem-solving instruction: Comparing students at risk of mathematics disability with and without risk of reading disability. *Journal of Learning Disabilities*, *37*, 293–306.

Gettinger, M., & Stoiber, K. (2008). Applying a response-to-intervention model for early literacy development in low-income children. *Topics in Early Childhood Special Education*, *27*, 198–213.

Good, R. H., & Shinn, M. R. (1990). Forecasting accuracy of slope estimates for reading curriculum-based measurement: Empirical evidence. *Behavioral Assessment*, *12*, 179–193.

Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*, *5*, 257–288.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial & Special Education*, *7*, 6–10.

Hartman, W. T., & Fay, T. A. (1996). *Cost-effectiveness of instructional support teams in Pennsylvania* (Policy Paper No. 9 of the Center for Special Education Finance). Palo Alto, CA: American Institutes for Research.

Hasbrouck, J. E., & Tindal, G. (1992). Curriculum-based oral reading fluency norms for students in Grades 2 through 5. *Teaching Exceptional Children*, *24*(3), 41–44.

Hasbrouck, J. E., & Tindal, G. (2005). *Oral reading fluency: 90 years of measurement* (Technical Report No. 33). Eugene, OR: Behavioral Research and Training.

Hintze, J. M., & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. *School Psychology Review*, *34*, 372–386.

Keller-Margolis, M. A., Shapiro, E. S., & Hintze, J. M. (2008). Long term diagnostic accuracy of curriculum-based measures in reading and mathematics. *School Psychology Review*, *37*, 374–390.

Marston, D., Muyskens, P., Lau, M., & Canter, A. (2003). Problem-solving model for decision-making with high-incidence disabilities: The Minneapolis experience. *Learning Disabilities Research and Practice*, *18*, 187–200.

McGlinchy, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review*, *33*, 193–203.

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children*, *71*, 445–463.

McNamara, K., & Hollinger, C. (2003). Intervention-based assessment: Evaluation rates and eligibility findings. *Exceptional Children, 69,* 181–193.

Pearson Education, Inc. (2008). AIMSweb. Retrieved June 21, 2008, from http://www.aimsweb.com

Peterson, D. W., Prasse, D. P., Shinn, M. R., & Swerdlik, M. E. (2007). The Illinois flexible service delivery model: A problem-solving model initiative. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 300–318). New York: Springer.

Shapiro, E. S., Keller, M. A., Lutz, G., Santoro, L., & Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment*, *42*, 19–35.

Shapiro, E. S., Solari, E., & Petscher, J. (2008). Use of a measure of reading comprehension to enhance prediction on the state high stakes assessment. *Learning and Individual Differences*, *18*, 316–328.

Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review*, *21*, 459–479.

Silberglitt, B., Burns, M. K., Madyun, N. H., & Lali, K. E. (2006). Relationship of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in the Schools*, *43*, 527–535.

Silberglitt, B., & Hintze, J. M. (2007). How much growth can we expect? A conditional analysis of R-CBM growth rates by level of performance. *Exceptional Children*, *74*, 71–84.

Speece, D. L., & Case, L. (2001). Classification in context: An alternative to identifying early reading disability. *Journal of Educational Psychology*, *93*, 735–749.

Stage, S. A., & Jacobsen, M. D. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review*, *30*, 407–419.

Success for All Foundation. (2008). *4sight benchmark assessments*. Baltimore: Author.

Sugai, G., & Horner, R. H. (2002). Introduction to the special series on positive behavior support in schools. *Journal of Emotional and Behavioral Disorders, 10,* 130–135.

VanDerHeyden, A. M., Witt, J. C., & Barnett, D. W. (2005). The emergence and possible futures of response to intervention. *Journal of Psychoeducational Assessment*, *23*, 339–361.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. *Journal of School Psychology*, *45*, 225–256.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional Children*, *69*, 391–409.

Vaughn, S., Wanzek, J., Linan-Thompson, S., & Murray, C. S. (2007). Monitoring response to supplemental services for students at risk for reading difficulties: High and low responders. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 234a–243). New York, NY: Springer.

Vaughn, S., Wanzek, J., Woodruff, A. L., & Linan-Thompson, S. (2007). Prevention and early identification of students with reading disabilities. In D. Haager, J. Klingner, & S. Vaughn (Eds.),

Evidence-based reading practices for response to intervention (pp. 11–27). Baltimore: Paul H Brookes Publishing.

Walker, H. M., Horner, R. H., Sugai, G., Bullis, M., Sprague, J. R., Bricker, D., et al. (1996). Integrated approaches to preventing antisocial behavior patterns among school-age children and youth. *Journal of Emotional and Behavioral Disorders*, *4*, 194–209.

Edward S. Shapiro, PhD, is the director of the Center for Promoting Research to Practice and professor of school psychology at Lehigh University, Bethlehem, Pennsylvania. His research interests include examining outcomes of response to intervention implementation, curriculum-based assessment as predictors of high-stakes assessment, and the evaluation of emergent literacy in preschool children.

Nathan H. Clemens, MEd, is a doctoral student in school psychology at Lehigh University. His current research interests include curriculum-based measurement, early identification of achievement problems, and the application of tiered models of prevention and intervention.